SPRING Working Group
Internet Draft
Intended status: Standards Track
Expires: August 26, 2024

Y. Liu
China Mobile
C. Lin
New H3C Technologies
S. Peng
Huawei Technologies
G. Mishra
Verizon Inc.
Y. Qiu
New H3C Technologies

February 22, 2024

Flexible Candidate Path Selection of SR Policy
draft-liu-spring-sr-policy-flexible-path-selection-05

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on August 26 2024.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

This document proposes a flexible SR policy candidate path selection method. Based on the real-time resource usage and forwarding quality of candidate paths, the head node can perform dynamic path switching among multiple candidate paths in the SR policy.

Table of Contents

1. Introduction

Segment routing (SR) [RFC8402] is a source routing paradigm that explicitly indicates the forwarding path for packets at the ingress node. The ingress node steers packets into a specific path according to the Segment Routing Policy (SR Policy) as defined in [RFC9256].

An SR Policy may have multiple candidate paths that are provisioned or signaled [I-D.ietf-idr-sr-policy-safi] [RFC8664] from one of more sources. The tie-breaking rules defined in [RFC9256] result in determination of a single "active path" in a formal definition.

Refer to [RFC9256] only the active candidate path MUST be used for forwarding traffic that is being steered onto that policy except for certain scenarios such as fast reroute where a backup candidate path may be used. A candidate path can be represented as a segment list or a set of segment lists. If a set of segment lists is associated with the active path of the policy, then the steering is per flow and weighted-ECMP (W-ECMP) based according to the relative weight of each valid segment list.

According to the criteria for the validity of candidate paths described in Section 5 of [RFC9256], if there is a valid segment list in the active candidate path, the active candidate path is still valid. When some segment lists of the active candidate path are invalid, the active candidate path may still be valid, but it may not continue to meet the actual forwarding requirements.

This document proposes a flexible SR policy candidate path selection method. Based on the real-time resource usage and forwarding quality of candidate paths, the head node can perform dynamic path switching among multiple candidate paths in the SR policy.

2. Terminology

The definitions of the basic terms are identical to those found in Segment Routing Policy Architecture [RFC9256].

3. Background of requirements

When some segment lists of the active candidate path are invalid, according to [RFC9256], if there is a valid segment list in the active candidate path, the active candidate path is still valid. But the paths of remaining segment lists may not meet the SR policy forwarding performance requirements, such as insufficient path bandwidth. Even if there are other candidate paths with lower preference that can meet the forwarding performance requirements in

the SR policy, the traffic will continue to be forwarded along the original active candidate path.

Take the following SR Policy as an example to explain in detail the problems existing in the current candidate path selection process.

```
SR Policy POL1
   Candidate Path CP1
      Preference 200
      Segment List 1 <SID11...SID1i>, Weight 1
      Segment List 2 <SID21...SID2j>, Weight 1
      Segment List 3 <SID31...SID3k>, Weight 1
   Candidate Path CP2
    Preference 100
      Segment List 4 <SID41...SID4i>, Weight 1
      Segment List 5 <SID51...SID5j>, Weight 1
      Segment List 6 <SID61...SID6k>, Weight 1
```

There are two static candidate paths CP1 and CP2 in SR policy POL1. CP1 has a higher preference. Both candidate paths are composed of three static segment lists with the same weight. The path indicated by each segment list can carry traffic of 100M bandwidth. When all Segment Lists in CP1 are valid, the candidate path can carry traffic with bandwidth less than 300M.

The bandwidth of the actual traffic forwarded by the SR policy is between 100M and 150M. Because the traffic forwarded on the candidate path will share the load on the three segment list paths according to the weight value. Therefore, normally, the candidate path can meet the forwarding requirements. The traffic is forwarded on the three segment lists of the high preference candidate paths of the SR policy.

When the segment list 1 and 2 in the high-preference candidate path CP1 are invalid, according to the candidate path validity criteria described in [RFC9256] Section 5, because the segment list 3 in CP1 is still valid, the active candidate path CP1 is still valid. All traffic of SR policy POL1 will continue to be forwarded through the path of CP1. However, because segment list 3 can only forward 100M traffic, over-bandwidth traffic will be discarded.

Of course, when the Segment List path fault is detected, the network device can report the detected fault information to the controller. The controller optimizes the forwarding path after receiving the message. However, this interaction process is relatively long, and it is difficult to meet the requirement for fast switching.

When the quality of high preference candidate paths deteriorates, such as insufficient available bandwidth, increased end-to-end transmission delay, and available segment lists that cannot meet service requirements, the same requirement exists. We hope to switch traffic to other candidate paths in the SR policy that better meet the forwarding quality requirements.

To solve this problem, this document proposes a new candidate path selection rule, which sets resource thresholds and forwarding quality requirements for candidate path. This candidate path can only be selected if the current path can meet the preset requirements.

4. Flexible Candidate Path Selection Method

As described in [RFC9256], the candidate path selection process operates primarily on the candidate path Preference. A candidate path is selected when it is valid and it has the highest Preference value among all the valid candidate paths of the SR Policy.

In the case of multiple valid candidate paths of the same Preference, the tie-breaking rules are evaluated on the identification tuple in the following order until only one valid best path is selected:

1. The higher value of Protocol-Origin is selected.

2. If specified by configuration, prefer the existing installed path.

3. The lower value of the Originator is selected.

4. Finally, the higher value of the Discriminator is selected.

This document proposes to take the forwarding quality requirements and resource requirements of candidate paths as the selection criteria of candidate paths.

Set the threshold parameters of forwarding quality and resources for candidate paths. First, find the paths that meet the threshold from the candidate paths of SR policy, and then select the best path as the active path according to the rules in the above standards.

Flexible candidate path selection method is more suitable for manually static configured SR policy paths. Unless otherwise specified, the candidate paths and segment lists mentioned in this document refer to static candidate paths and static segment lists.

4.1. Threshold Parameters of Candidate Paths

The threshold parameters of candidate paths can include but are not limited to the following:

* Jitter

* Latency

* Packet loss

  Delay, jitter, and packet loss are thresholds at the segment list level.

  When the jitter, delay, or packet loss of a valid segment list cannot meet the specified threshold requirement, the segment list will be treated as an invalid segment list and will no longer load share traffic.

* Available bandwidth

* The bandwidth threshold is the threshold at the candidate path level, which corresponds to the sum bandwidth of segment list in the candidate path.

  The available bandwidth refers to the sum of the preset bandwidth of all valid segment lists in the candidate path that meet the threshold requirements for latency, jitter, or packet loss.

  The available bandwidth is the sum of preset bandwidth of all valid segment lists in the candidate path, or the cumulative value calculated based on the weight and preset bandwidth of each segment list.

* Actual bandwidth

  The actual bandwidth refers to the sum of the actual available remaining bandwidth of each valid segment list in the candidate path.

  Due to the different congestion conditions of each node on the forwarding path, the actual bandwidth that can forward service packets may differ from the preset bandwidth. By utilizing some measurement mechanisms, the actual minimum available bandwidth and actual minimum remaining bandwidth of all nodes along the path can be obtained. The specific measurement mechanism is not within the scope of this document.

* Precision Availability Metrics (PAM)

  Consider a candidate path of SR policy as a Service Level Objective (SLO), based on the Precision Availability Metrics (PAM) defined in [I-D.ietf-ippm-pam], determine whether the candidate path meets the forwarding requirements.

If both segment list level thresholds (such as latency, jitter, or packet loss) and candidate path level thresholds (such as available bandwidth) are specified, when one or some segment lists in the candidate path do not meet the threshold of segment list level, it is necessary to continue checking the quality of the candidate path. If the quality of the candidate path still meets the requirement, traffic can continue to be forwarded along that candidate path.

For example, two threshold parameters, delay and available bandwidth, are specified simultaneously for the candidate path with multiple segment lists. When the delay of a segment list exceeds the threshold, the following processing is performed:

1. Remove the segment list from the forwarding path first.

2. Next, check if the total bandwidth of the remaining valid segment lists still meets the bandwidth threshold requirements.

   * If the bandwidth still meets the requirements, the path still meets the forwarding quality requirements, and the traffic is still forwarded along this path.

   * Otherwise, it should be considered that the path no longer meets the quality requirements.

If the candidate path does not specify any threshold parameters, select the primary candidate path according to the selection method defined in [RFC9256].

By default, there is no threshold parameter specified on the candidate path.

4.2. Rules for Selecting the Best Path

When the current forwarding quality and hardware resources of a candidate path meet the specified threshold requirements, it only means that this candidate path could forward traffic.

If there are multiple candidate paths in the SR policy that meet the forwarding requirements at the same time, the candidate paths need to be sorted to select the best one.

Under the condition that multiple valid candidate paths meet the threshold requirements, evaluate the tie breaking rules in the following order until only one valid best path is selected:

1. If the quality requirements of the candidate path are specified, it is necessary to check whether the path meets the quality requirements. Only the valid path that meets the quality requirements can be selected as the active path.

   If only one path in the SR policy meets the quality requirements, the path is selected.

   If multiple candidate paths meet the quality requirements at the same time, or if all candidate paths fail to meet the requirements, then select the following second step according to the Preference.

2. The higher value of the Preference is selected.

3. The higher value of Protocol-Origin is selected.

4. If specified by configuration, prefer the existing installed path.

5. The lower value of the Originator is selected.

6. Finally, the higher value of the Discriminator is selected.

4.3. Flexible Candidate Path Selection Process

The process of selecting the best path for SR policy through the threshold parameter of the path is as follows.

1. Configure the threshold parameters on the candidate path of the head node through static manual configuration or controller distribution.

2. The head node monitors whether the available resources and forwarding quality of the SR policy candidate path exceed the thresholds.

3. The forwarding quality of path can be obtained through active or passive performance measurement methods, such as iOAM [RFC9378], STAMP [I-D.ietf-spring-stamp-srpm], TWAMP [RFC5375], etc. The real-time quality data can be calculated by the controller and distributed to the head node, or calculated by the head node according to the network measurement data. The measurement method

and quality data acquisition method are beyond the scope of this document.

4. According to the rules described in Section 4.2, when the available resources are less than the threshold, or the forwarding quality cannot meet the threshold requirements, select a new active candidate path.

5. After the old active candidate path eliminates the fault or improves the forwarding quality, whether to recover can be specified by the configuration. If fault recovery is required, start a wait timer for delay recovery. When the timer expires and the old active candidate path still meets the threshold requirements, the traffic will be switched to the old higher preference candidate path.

For avoiding path switching frequently, both over-threshold switching and fault recovery should be delayed. The interval of delay waiting can be adjusted by configuration.

In order to distribute the threshold parameters of SR Policy to the head node, it is necessary to extend the control plane, such as NetConf, PCEP and BGP. The extensions of BGP and PCEP are described in [I-D.liu-idr-bgp-sr-policy-cp-threshold] and [I-D.liu-pce-sr-policy-cp-threshold].

5. Usecases of Flexible Candidate Path Selection

The SR policy in Section 3 is still used to illustrate how the flexible candidate path selection method switches candidate paths.

SR policy POL1 has two candidate paths CP1 and CP2. The Preference of CP1 is 200, and the Preference of CP2 is 100. Both candidate paths are composed of three segment lists with the same weight.

5.1. Select the Best Path Based on End-to-End Delay

The quality requirement for the services carried on the SR policy is that the transmission delay must be less than 200ms. The bandwidth of the actual traffic forwarded by the SR policy is between 100Mbps and 150Mbps.

When the delay of Segment List 1 does not meet the requirements, continue to check the available bandwidth of CP1. Due to segment list 2 only having 100Mbps bandwidth, it cannot meet the actual traffic forwarding requirements. CP2 is selected as the new active candidate path of POL1. The traffic forwarded by POL1 is switched to the path of CP2 for forwarding.

```
SR Policy POL1
  Candidate Path CP1
     Preference 200
     Delay threshold 200ms //Delay<=200ms
     Segment List 1 <SID11...SID1i>, Weight 1 //100M, Delay>1s
     Segment List 2 <SID21...SID2i>, Weight 1 //100M, Delay<100ms
  Candidate Path CP2
     Preference 100
     Delay threshold 200ms  //Delay<=200ms
     Segment List 3 <SID31...SID3i>, Weight 1 //100M, Delay<100ms
     Segment List 4 <SID41...SID4i>, Weight 1 //100M, Delay<100ms
```

5.2. Select the Best Path Based on Available Bandwidth

The path indicated by each segment list can carry traffic of 100Mbps bandwidth. When the Segment Lists are valid, the candidate path can carry traffic with bandwidth less than 300Mbps. The bandwidth of the actual traffic forwarded by the SR policy is between 100Mbps and 150Mbps.

```
SR Policy POL1
   Candidate Path CP1
      Preference 200
      Available bandwidth ratio 50
      Segment List 1 <SID11...SID1i>, Weight 1
      Segment List 2 <SID21...SID2j>, Weight 1
      Segment List 3 <SID31...SID3k>, Weight 1
   Candidate Path CP2
    Preference 100
    Available bandwidth ratio 50
      Segment List 4 <SID41...SID4i>, Weight 1
      Segment List 5 <SID51...SID5j>, Weight 1
      Segment List 6 <SID61...SID6k>, Weight 1
```

First, take the available bandwidth as the threshold parameter of POL1. The threshold for configuring the ratio of available bandwidth is 50%. When the available bandwidth of the candidate path is less than 50%, path switching is performed.

Normally, the three segment lists of CP1 and CP2 are valid. The available bandwidth of CP1 is 300M, and the ratio of available bandwidth is 100%, which can meet the threshold requirements of the path. So CP1 is selected as the active candidate path according to the Preference.

If the paths indicated by Segment 1 and 2 fail, Segment List 1 and 2 become invalid, and the available bandwidth of CP1 becomes 100M. The ratio of available bandwidth becomes 33.3% (i.e., 100/300). Because

the ratio of available bandwidth of CP1 is lower than the specified

threshold, CP1 has failed to meet the forwarding quality requirements. Need to reselect the active candidate path for POL1.

The three segment lists of the low-preference candidate path CP2 of POL1 are valid, and the available bandwidth can meet the threshold requirements. CP2 is selected as the new active candidate path of POL1. The traffic forwarded by POL1 is switched to the path of CP2 for forwarding.

5.3. Select the Best Path Based on Actual Bandwidth

The quality requirement for the services carried on the SR policy is that the actual available bandwidth of the forwarding path must be greater than 80Mbps. When there is traffic congestion on a node in the Segment List 1 path, a maximum of 50Mbps service traffic can be forwarded. If Segment List 2 is in a down state or the delay has exceeded the threshold, the path of Segment List 2 will not load share traffic.

Because the sum of the actual bandwidth of CP1 is less than 80Mbps, CP2 will be selected as the new active candidate path of POL1. The traffic forwarded by POL1 is switched to the path of CP2 for forwarding.

```
   SR Policy POL1
     Candidate Path CP1
        Preference 200
        Remaining bandwidth 50mbps
         Segment List 1 <SID11...SID1i>, Weight 1
                    //bandwidth of 100M, actual remaining 30M
         Segment List 2 <SID21...SID2j>, Weight 1
                    //In Down state, or the delay has exceeded the
threshold.
     Candidate Path CP2
        Preference 100
        Remaining bandwidth 50mbps
         Segment List 3 <SID41...SID4i>, Weight 1 //100M
         Segment List 4 <SID51...SID5j>, Weight 1 //100M
         Segment List 5 <SID61...SID6k>, Weight 1 //100M
```

6. IANA Considerations

This document has no IANA actions.

7. Security Considerations

This document does not introduce any security considerations.

8. References

8.1. Normative References

[I-D.ietf-idr-sr-policy-safi] Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and Jain, D., "Advertising Segment Routing Policies in BGP", draft-ietf-idr-sr-policy-safi-00 (work in progress), February 2024.

[I-D.ietf-spring-stamp-srpm] Gandhi, R., Filsfils, C., Voyer, D., Chen, M., Foote, R., "Performance Measurement Using Simple TWAMP (STAMP) for Segment Routing", draft-ietf-spring-stamp-srpm-11 (work in progress), February 2024.

[I-D.liu-idr-bgp-sr-policy-cp-threshold] Liu, Y., Lin, C., Qiu, Y., " BGP Extension for Distributing CP Threshold Constraints of SR Policy", draft-liu-idr-bgp-sr-policy-cp-threshold-00 (work in progress), October 2023.

[I-D.liu-pce-sr-policy-cp-threshold] Liu, Y., Lin, C., Qiu, Y., " PCEP Extension to Support Signaling Candidate Path Threshold Constraints of SR Policy", draft-liu-pce-sr-policy-cp-threshold-00 (work in progress), October 2023.

[I-D.ietf-ippm-pam] Mirsky, G., Halpern, J., Min, X., Clemm, A., Strassner, J., Francois, J., "Precision Availability Metrics for Services Governed by Service Level Objectives (SLOs)", draft-ietf-ippm-pam-09 (work in progress), December 2023.

[I-D.Ravi-app-csig] Ravi, A., Dukkipati, N., Mehta, N., Kumar, J., "Congestion Signaling (CSIG)", draft-ravi-ippm-csig-01 (work in progress), February 2024.

[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>.

[RFC5375] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., Babiarz, J., "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5375, DOI 10.17487/RFC5375, October 2008, <https://www.rfc-editor.org/info/rfc5375>.

[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>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., Hardwick, J., "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[RFC9256] Filsfils, C., Talaulikar, K., 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>.

[RFC9378] Brockners, F., Bhandari, S., Bernier, D., Mizrahi, T., "In Situ Operations, Administration, and Maintenance (IOAM) Deployment", RFC 9378, DOI 10.17487/RFC9378, April 2023, <https://www.rfc-editor.org/info/rfc9378>.

8.2. Informative References

TBD

9. Acknowledgments

The authors would like to thank the following for their valuable contributions of this document:

TBD

Authors' Addresses

Yisong Liu
China Mobile
Beijing
China

Email: liuyisong@chinamobile.com

Changwang Lin
New H3C Technologies
Beijing
China

Email: linchangwang.04414@h3c.com

Shuping Peng
Huawei Technologies
Beijing
China
Email: pengshuping@huawei.com

Gyan S. Mishra
Verizon Inc.
Email: gyan.s.mishra@verizon.com

Yuanxiang Qiu
New H3C Technologies
Beijing
China

Email: qiuyuanxiang@h3c.com