

Workgroup: Spring Working Group
Internet-Draft:
draft-deng-spring-sr-loop-free-01

Published: 23 November 2023

Intended Status: Informational

Expires: 26 May 2024

Authors: L. Deng Y. Zhu X. Geng
 China Telecom China Telecom Huawei Technologies
 Z. Hu
 Huawei Technologies

SR based Loop-free implementation

Abstract

Microloops are brief packet loops that occur in the network following a topology change (link down, link up, node fault, or metric change events). Microloops are caused by the non-simultaneous convergence of different nodes in the network. If nodes converge and send traffic to a neighbor node that has not converged yet, traffic may be looped between these two nodes, resulting in packet loss, jitter, and out-of-order packets. This document presents some optional implementation methods aimed at providing loop avoidance in the case of IGP network convergence event in different scenarios.

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 26 May 2024.

Copyright Notice

Copyright (c) 2023 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

- [1. Introduction](#)
 - [2. Conventions used in this document](#)
 - [3. Anti-Microloop Scheme for Switching Scenarios](#)
 - [4. Anti-Microloop Scheme for Back-switching Scenarios](#)
 - [5. Anti-Microloop Scheme for Multi-source Scenarios](#)
 - [6. Anti-Microloop Scheme for Multi-point Scenarios](#)
 - [7. Conclusion](#)
 - [8. Security Considerations](#)
 - [9. IANA Considerations](#)
 - [10. Acknowledgement](#)
 - [11. Normative References](#)
- [Authors' Addresses](#)

1. Introduction

An IP network computes paths based on the distributed IGP protocols. If a node or link fails, a loop may occur on the network because LSDBs are not synchronized. Take the IS-IS/OSPF link-state protocol as an example. Each time the network topology changes, some routers need to update the FIB table based on the new topology. Due to the different convergence time and convergence orders, different routers may be asynchronous for a short time. Depending on the capability, configuration parameters, and service volume of the device, the database may not be synchronized in milliseconds to seconds. During this period, each device on the packet forwarding path may be in the pre-convergence state or the post-convergence state. If the status is not synchronized, forwarding routes may be inconsistent and a forwarding loop may occur. However, such a loop disappears after all devices on the forwarding path complete convergence. Such a transient loop is called a "microloop". Microloops may cause packet loss, delay variation, and packet disorder on the network.

The Segment Routing defined in [[RFC8042](#)] . can be used to cope with microloop issue on the network. When a loop may occur due to a network topology change, a network node creates a loop-free segment list to direct traffic to the destination address. After all network nodes converge, the network node returns to the normal forwarding state. This effectively eliminates loops on the network.

[\[I-D.bashandy-rtgwg-segment-routing-uloop\]](#) describes the basic principles of how to use Segment Routing to cope with microloop. This document describes some optional implementation methods of SR for microloop avoidance in different scenarios.

2. Conventions used in this document

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 [\[RFC2119\]](#) .

3. Anti-Microloop Scheme for Switching Scenarios

Switching microloops refer to the microloop caused by node/link failures. Along the traffic forwarding path, a loop may be caused if a node closer to the point of failure converges before a node far from the point of failure. Figure 1 is used as an example to describe the switching microloop caused process: when the link between R3 and R5 fails, it is assumed that R3 completes convergence first and R2 does not complete convergence. R1 and R2 forward the packet along the previous path to R3. Since R3 has converged, it forwards the traffic to R2 according to the route after convergence. Thus, the switching microloop happens between R2 and R3.

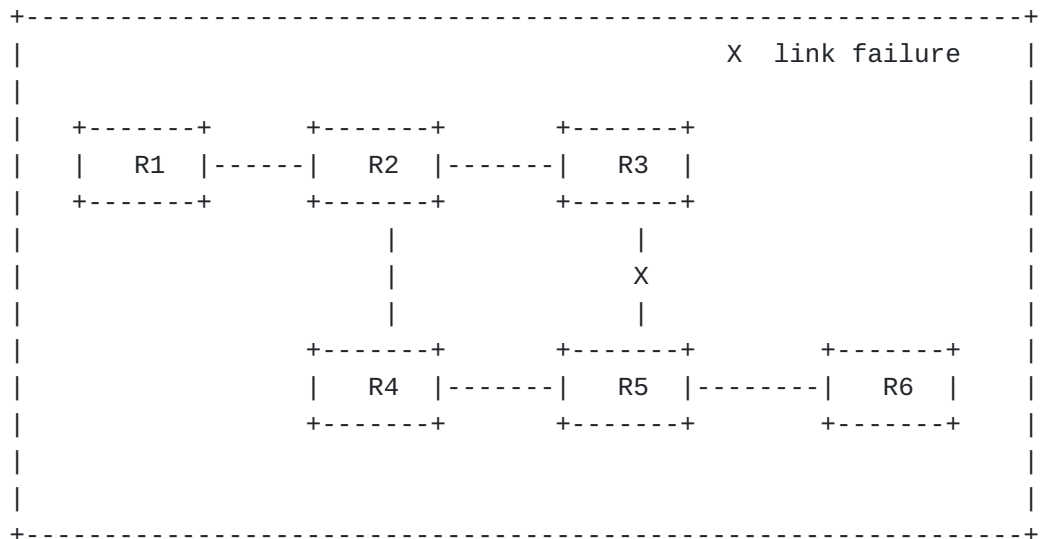


Figure 1: Switching illustrative scenario, failure of link R3-R5

TI-LFA is deployed in all nodes of the network, and when the link between R3 and R5 fails, the convergence process after deploying switching anti-microloop is as follows:

*Phase 1: A hold-down timer T1 is configured on R3 (R3 is the neighboring node of the failed node/link) and R3 uses TI-LFA forwarding for the duration of T1;

*Phase 2: A hold-down timer T2 is configured on the remote node and the node forwards traffic to R3 (specify the Node Sid of R3) for the duration of T2;

*Phase 3: T2 timeout, the remote node returns to normal convergence firstly;

*Phase 4: T1 timeout, R3 reverts back to normal convergence.

Time T1 must be longer than time T2. This scheme is limited to single point of failure, the TI-LFA backup path may be affected in case of multi-point failure.

4. Anti-Microloop Scheme for Back-switching Scenarios

Microloops may occur not only when the node/link fails, but also after the failure node/link recovering. Figure 2 is used as an example to introduce the process of the back-switching microloop. After the failure node/link recovering, a loop may be caused if a node further from the point of failure converges before a node closer to the point of failure.

R1 forwards the traffic to the destination node R6 following the path R1->R2->R3->R5->R6. When the link between R2 and R3 fails, R1 forwards the traffic to the destination node R6 following the re-converged path R1->R2->R4->R5->R6. After the failure link between R2 and R3 is recovered, assuming that R4 is the first to complete convergence, R1 forwards the traffic to R2. Since R2 has not completed convergence, the packet is still forwarded to R4 in accordance with the path before the failure link recovering. R4 has already completed convergence, so R4 forwards it to R2 in accordance with the path after the failure link recovering, and the microloop occurred between R2 and R4.

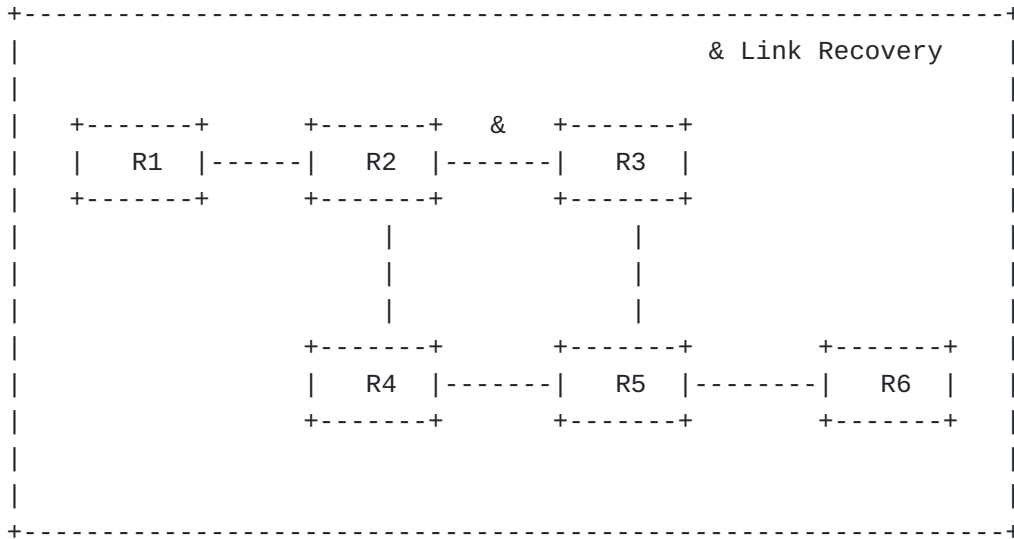


Figure 2: Back-switching illustrative scenario, recovery of link R2-R3

Since the network does not enter the TI-LFA forwarding process after the node/link failure is recovered, the delay convergence cannot be used in the back-switching scenario to prevent the generation of microloops as in the switching scenario. In the back-switching scenario, we only need to specify the Adj-SID of the back-switching link to achieve loop-free.

From the above process of back-switching microloop generation, it can be seen that microloops happens because R4 is unable to pre-install a loop-free path computed for link up. Therefore, in order to eliminate potential loop after the the faulty node/link recovering, R4 needs to be able to converge to a loop-free path.

When the faulty node/link is recovered, the path can be anti-microloop by simply specifying Adj-SIDs of the neighbor node. As shown in Figure. 2, R4 senses that the faulty link R2-R3 is recovered and re-converges to the destination R6 with the R4->R2->R3->R5->R6 path. The recovery of the faulty link R2-R3 does not affect the SR path from R4 to R2, so the path from R4 to R2 must be a loop-free path. Similarly, the path from R3 to R6 is not affected by the recovery of the failed R2-R3 link, and the path from R3 to R6 must be loop-free. The only thing affected is the path from R2 to R3. The loop-free path from R4 to R6 can be determined by just specifying the path from R2 to R3. So it is only necessary to insert an End.X SID from R2 to R3 in the converged path of R4 End. X SID instructs the message to be forwarded from R2 to R3, and the path from R4 to R6 is guaranteed to be loop-free.

5. Anti-Microloop Scheme for Multi-source Scenarios

When an IPv4 or IPv6 prefix is advertised by multiple nodes in an IS-IS domain, the prefix has multiple route sources, which is called

a multi-source route. This section is for the multi-source microloop avoidance scenario, which may occur when multiple nodes advertise the same route with inconsistent convergence speeds.

SRv6 multi-source microloop prevention mainly uses SRv6 END.X and END SID as the label stack for multi-source microloop prevention. SR-MPLS mainly uses the prefix SID and Adj SID as the label stack for multi-source anti-microloop.

The following example is to describe how microloop happens when multiple nodes advertise the same route.

1. R3 and R6 both import the route 2001:db8:3::. The link between R2 and R3 fails. It is assumed that R2 first completes convergence, and R1 hasn't completed convergence yet.
2. R1 forwards the packet to R2 along the path before the failure.
3. Because R2 has completed convergence, R2 forwards packets to R1 according to the next hop of the route. In this way, a loop is formed between R1 and R2.

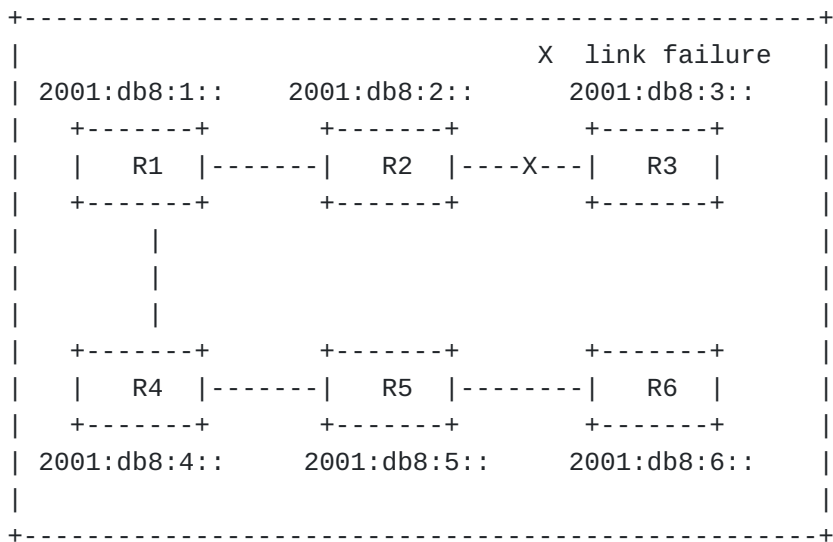


Figure 3: Multi-source illustrative scenario, failure of link R2-R3

A possible solution is that: the preferred destination node of the packets destined for 2001:db8:3:: changes from R3 to R6, but the convergence path from R2 to R5 does not change. In this case, timer T1 on R2 can be started. Before T1 expires, for a packet that accesses the R6, an End.X SID between the R5 and the R6 or an End SID of the R6 is added to the encapsulation in order to ensure that the packet is forwarded to the R6. A basic principle is similar to that of SR-MPLS.

6. Anti-Microloop Scheme for Multi-point Scenarios

TBD

7. Conclusion

There are various scenarios and different implementation methods for loop prevention. The implementation methods proposed by this document based on SR microloop avoidance mechanism can be used for subsequent research and development.

8. Security Considerations

The behavior described in this document is internal functionality to a router that result in the ability to explicitly steer traffic over the post convergence path after a remote topology change in a manner that guarantees loop freeness. Because the behavior serves to minimize the disruption associated with a topology changes, it can be seen as a modest security enhancement.

9. IANA Considerations

No requirements for IANA.

10. Acknowledgement

The authors would like to thank everyone who contributed to the draft.

11. Normative References

[I-D.bashandy-rtgwg-segment-routing-uloop]

Bashandy, A., Filsfils, C., Litkowski, S., Decraene, B., Francois, P., and P. Psenak, "Loop avoidance using Segment Routing", Work in Progress, Internet-Draft, draft-bashandy-rtgwg-segment-routing-uloop-15, 18 June 2023, <<https://datatracker.ietf.org/doc/html/draft-bashandy-rtgwg-segment-routing-uloop-15>>.

[I-D.ietf-rtgwg-segment-routing-ti-lfa]

Litkowski, S., Bashandy, A., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-12, 17 November 2023, <<https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-segment-routing-ti-lfa-12>>.

[I-D.ietf-spring-segment-protection-sr-te-paths]

Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu, "Segment Protection for SR-TE Paths", Work in Progress,

Internet-Draft, draft-ietf-spring-segment-protection-sr-te-paths-05, 27 September 2023, <<https://datatracker.ietf.org/doc/html/draft-ietf-spring-segment-protection-sr-te-paths-05>>.

[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>>.

[RFC8042] Zhang, Z., Wang, L., and A. Lindem, "OSPF Two-Part Metric", RFC 8042, DOI 10.17487/RFC8042, December 2016, <<https://www.rfc-editor.org/info/rfc8042>>.

Authors' Addresses

Lijie Deng
China Telecom
109, West Zhongshan Road, Tianhe District
Guangzhou
Guangzhou, 510000
China

Email: denglj4@chinatelecom.cn

Yongqing Zhu
China Telecom
109, West Zhongshan Road, Tianhe District
Guangzhou
Guangzhou, 510000
China

Email: zhuyq8@chinatelecom.cn

Xuesong Geng
Huawei Technologies
Huawei Building, No.156 Beiqing Rd
Beijing
Beijing, 100095
China

Email: gengxuesong@huawei.com

Zhibo Hu
Huawei Technologies
Huawei Building, No.156 Beiqing Rd
Beijing
Beijing, 100095
China

Email: huzhibo@huawei.com